

Discourse markers and registers, a corpus-based analysis

Tora Hedin
(Stockholm University)

pragmatics, translation, parallel corpora, Czech, discourse markers

The present study analyzes discourse markers in three different registers in Czech and Swedish: texts of fiction, spoken political texts and legislative texts. While corpus-based, it employs both quantitative and qualitative methods.

Research questions:

- How are the selected Czech discourse markers used in the three registers?
- Which are the translation correlates in Swedish in the three registers?
- How can translation corpora help understanding the semantics of the discourse particles?

Contrastive analysis, for example Swedish-English, has been widely used by researchers in the analysis of discourse markers. (Aijmer has published numerous studies of discourse markers in Swedish and English: Aijmer 2007, Aijmer and Simon-Vandenberg 2004). In a Czech context, publications have started to emerge in recent years (Šindlerová and Štěpánková 2021, Štěpánková et al. 2023, Šindlerová et al. 2023).

The research design for this study comprises selecting particles (Čermák & Křen 2011) and establishing their dictionary meanings. Their usage is then compared in three different registers: fiction, spoken political texts and legislative texts. The subsequent step is to analyze their Swedish translation parallels in the three registers. The empirical data was collected from the InterCorp14 Czech¹ and InterCorp14 Swedish: texts of fiction (Core), transcripts from proceedings from the European parliament (EUROPARL; Koehn, 2005) and ACQUIS, legislative EU texts. In this paper, one discourse marker has been selected: *právě* (*just, right, exactly*). In two of the subcorpora, Czech and Swedish are source language (SL), while in the third (ACQUIS) both Czech and Swedish are translated.

The results of the analyses indicate that the Czech particle *právě* indeed is used in different ways in the three registers. Tentative results show that it differs as regards to function in the Czech texts, as well as in the number of Swedish equivalents and in the translation strategies employed in the three registers: In FICTION there are a large number translation parallels, but one of them, *just*, is fairly dominant. In EUROPARL, *právě* is often omitted, changing the meaning of the utterance.

¹ korpus.cz: <https://wiki.korpus.cz/doku.php/en:cnk:intercorp:verze14>

The translation of Czech particles into Swedish is complex with lexical/grammatical translations and frequent omissions. These findings can prove useful for such purposes as language comparison, translation, bilingual lexicography and L2 teaching.

References

- Aijmer, Karin. 2007. The meaning and functions of the Swedish discourse marker *alltså* – Evidence from translation corpora. *Catalan Journal of Linguistics* 6, 2007 31-59.
- Aijmer, Karin and Anne-Marie Simon-Vandenberg. 2004. A model and a methodology for the study of pragmatic markers: the semantic field of expectation. *Journal of Pragmatics*, 36(10), 1781–1805.
- Čermák, František, and Michal Křen. 2011. *A Frequency Dictionary of Czech*. Routledge Frequency Dictionaries. Oxon: Routledge.
- Koehn, Philipp. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In: Hutchins, J. (ed.) *The Tenth Machine Translation Summit Conference Proceedings*. [Online]. International Association for Machine Translation, pp. 79-86. Available at: <https://www.research.ed.ac.uk/en/publications/europarl-a-parallel-corpus-for-statistical-machine-translation>
- Šindlerová, Jana and Barbora Štěpánková. 2021. Between adverbs and particles: A corpus study of selected intensifiers. *Journal of Linguistics/Jazykovedný časopis*, 72(2), 444–453.
- Šindlerová, Jana, Barbora Štěpánková and Ingrid Andrén Lindgren. 2023. Epistemická částice zřejmě pohledem paralelního korpusu. *Korpus - Gramatika - Axiologie*, 27(2023), 37–52.
- Štěpánková, Barbora, Jana Šindlerová and Lucie Poláková. 2023. The epistemic marker *určitě* in the light of corpus data. *Jazykovedný Časopis*, 74(1), 130–139.